CAREER: Data-efficient task creation for generalized and interpretable image analysis
PI: Abdullah-Al-Zubaer Imran (University of Kentucky)

Deep learning models for computer vision are effective when large amounts of labeled data are available in supervised learning settings. Obtaining annotations can be challenging — time-consuming, laborious, error-prone, and subjective. Self-supervised representation learning (SSRL) and foundation models based on SSRL are a good alternative to supervised learning models as they enable learning with limited labeled data. However, they require a large pool of unlabeled data which can be challenging to obtain for some sensitive domains like medicine. There are many a number of different choices for self-supervised learning tasks built on computer vision domains which are also used in other domains like medical imaging. There is no clear guideline or path to automatically selecting the optimal self-supervised learning task for data in an arbitrary domain and for a given task information.

We, therefore, propose the **GUIDR (Generalizable, Understandable, and Interpretable Domain-specific Representations)** framework to enhance both generalization and interpretability in vision and multimodal AI systems. The core of GUIDR is a reasoning engine that matches the characteristics of input data, such as modality, structure, and task type, with optimal SSRL task strategies automatically created from a curated knowledge base. GUIDR supports both training from scratch and adaptation of existing foundation models, with modular pipelines for handling various image analysis tasks such as image classification, segmentation, object detection, and image quality assessment (IQA).

Intellectual Merit
The project introduces a novel fusion of SSRL, meta-reasoning, and human-in-the-loop learning to address two long-standing challenges in AI — brittle generalization and poor interpretability. By automatically creating pretext tasks and selecting model architectures based on domain-/task-specific priors, in Aim 1, GUIDR will produce representations that are both transferable and semantically aligned with expert knowledge. In Aim 2, GUIDR will first be evaluated on standard computer vision benchmarks to establish its ability to learn transferable, task-relevant representations. It will then be extended to high-impact, domain-specific applications in medicine, geoscience, and disaster response. In Aim 3, to ensure transparency and trustworthiness for all the critical applications, GUIDR will generate multimodal explanations linking visual outputs to natural language justifications, allowing users to interpret model decisions. The proposed research advances the science of automated representation learning by formalizing a domain-task-pretext mapping space (Aim 1), developing generalized reasoning modules to explore it (Aims 1-2), and validating outcomes across different real-world data domains (Aims 2-3), and enhancing the trustworthiness of model decisions (Aim 3).

Broader Impacts
This project is designed to deliver substantial broader impacts through a comprehensive approach to education, community engagement, and interdisciplinary science. The project supports an early-career computer science faculty at a major R1 institution in the ***EPSCoR jurisdiction***. Graduate and undergraduate students will be directly involved in innovative, hands-on research, preparing them as future leaders equipped with interdisciplinary expertise. The innovative AI models and the findings from GUIDR will be used to inform classroom teaching for courses taught by the PI. GUIDR will also focus on engaging K-12 students from a local magnet program and young children through the Early Learning Center toward increasing AI literacy aligned with the Executive Order on ***Advancing AI Education for American Youth***. GUIDR is expected to democratize access to high-performing, trustworthy AI for disciplines with limited data and high demands for transparency, such as medicine, disaster response, and environmental monitoring. By reducing reliance on large labeled datasets while maintaining interpretability, it empowers practitioners to build deployable models with confidence. In medicine, this may translate to more reliable diagnoses from limited scans, while in disaster response, it can support rapid and explainable damage assessment. The project will also contribute open-source code, curated domain-specific SSRL strategies, and educational resources to accelerate learning and adoption in both academic and applied settings. Overall, it promotes ethical, interpretable, and generalizable image analysis systems that adapt to real-world needs.